New Implementation of and the Modeling by the Extended Simulated Annealing Process to Structures of T4 Lysozyme Mutants at the 86th Residue

SHIGERU ENDO,* JUNICHI HIGO,† and KUNIAKI NAGAYAMA
Protein Array Project, ERATO, JRDC, 5-9-1, Tokodai, Tsukuba 300-26, Japan

HIROSHI WAKO
School of Social Sciences, Waseda University, Nishi-Waseda, Shinjuku-ku, Tokyo 169-50, Japan

Received 4 January 1995; accepted 20 June 1995

ABSTRACT

The extended simulated annealing process (ESAP) is a useful method for modeling the partial structure of proteins [J. Higo et al., Biopolymers, **32**, 33 (1992)]. In ESAP, a protein molecule is divided into two parts: small, flexible fragments constituting the concerned partial structure, and the remaining part, for which the structure is kept rigid during the simulation. We have improved the program of ESAP so that it can be adapted to general macromolecules. Any sidechain on the rigid part can be set to rotate. Soft repulsion between van der Waals spheres is introduced to avoid conformational trapping into local minima. This improved program was tested for modeling structural changes caused by eight kinds of amino acid mutation at the 86th residue in T4 lysozyme. For each mutant we obtained a model structure that was close to the X-ray structure. The root mean square (rms) deviations from the X-ray structure were 0.3 to 0.8 Å for all heavy atoms and about 0.2 Å for the main-chain atoms. We also modeled the structure of an Ile mutant, for which the X-ray structure has not yet been reported. ESAP can be used to model structural changes due to a single residue mutation in proteins.

*Author to whom all correspondence should be addressed at Department of Physics, School of Science, Kitasato University, Kitasato, Sagamihara 228, Japan.
†Current address: School of Pharmaceutical Science, Kitasato University, Shirokane 5, Minato-Ku, Tokyo 108, Japan.

Journal of Computational Chemistry, Vol. 17, No. 4, 476–488 (1996)

CCC 0192-8651 / 96 / 040476-13

Introduction

Protein molecules are composed of thousands of atoms. A large number of degrees of freedom are needed to describe the conformation of the protein. Thus it is difficult to explore the whole conformational space, even with the fastest computer currently available. A simple way to avoid this difficulty is to reduce the degrees of freedom. When a theme related only to a partial structure of a protein is to be investigated by computer simulation, a limited part of the protein should be treated as flexible and the rest as rigid or constrained to the known structure. Based on this computational framework, the extended simulated annealing process (ESAP) was developed[1] and applied to the modeling of hypervariable regions of immunoglobulins[1,2] and loops of trypsin and trypsin inhibitor.[3]

In ESAP, a protein molecule is divided into two parts: a flexible part consisting of one or more flexible fragments, and a rigid part. Dihedral angles are used as independent variables to describe the molecular conformations of the flexible fragments. All dihedral angles in the rigid part are fixed. The N-termini of the flexible fragments are also fixed to the rigid part. Simulated annealing has been done by efficient Monte Carlo simulations, which uses scaled collective variables.[4] The structures obtained are average conformations at thermodynamic equilibrium. Besides ESAP, a similar method that numerically solved the chain closure problem has been developed to model a partial structure of proteins.[5]

A variety of computer programs that treat proteins based on their dihedral angles have been developed.[6-9] These programs rapidly and analytically calculate an ECEPP energy function,[10,11] its derivative, and its Hessian. We had previously extended the computer program FEDER (Fast Ecepp DERivative)[8] so that it can be widely used for macromolecules as well as proteins.[12] Recently, FEDER was improved to be adapted to molecules of multiple chains.[13] This improved program, called FEDER/2, can rapidly make analytical calculations of the Hessian with respect to six translation–rotation variables as well as dihedral angles. From the viewpoint of program structure, FEDER/2 is suitable to implement ESAP. In the improved ESAP implemented with FEDER/2, the N-termini of the flexible fragments are not fixed to the rigid part but are left free to move around it. Other improvements have also been added as follows: (1) Any sidechain on the rigid part can be set to rotate, and such sidechains are included in the flexible part; (2) the method to calculate the Hessian, which was used to obtain the scaled collective variables,[4] was changed from a numerical method to a fast analytical one; (3) the energy parameters used were changed from those for united atoms by Robson and Platt[14] to those of ECEPP/2[10,11]; and (4) soft repulsion between van der Waals spheres was introduced to avoid conformational trapping into local minima.

Conformational changes caused by single residue mutation were normally localized around the mutated residue. Modeling of local structures around mutated residues has been carried out using energy minimization,[15] sidechain optimization by packing criteria,[16,17] and a modified method of the free energy perturbation approach.[18] In this study we applied the ESAP method implemented with FEDER/2 to the modeling of structural changes caused by mutation at the 86th residue in T4 lysozyme, where Pro86 in the wild type was substituted with 10 different amino acid residues.[19] The model structures for eight mutants, for which the crystal structures were already solved with X-ray crystallography,[19,20] were obtained with annealing processes starting from an extended conformation. These were compared with the X-ray structures. We also modeled the structure of the Ile mutant, for which the X-ray structure has not yet been reported.

Methods

TREATMENT OF THE MOLECULAR SYSTEM

In this article we consider structural changes of a protein by single-residue mutation. A diagram of the molecular system, which is a general scheme to model a partial structure of a protein, is shown in Figure 1. The protein was divided into two parts: One is a flexible part that included the mutated residue and surrounding residues, and the other is the rigid part. We set the structure of the rigid part so that it remained unchanged after residue mutation. The flexible part can be composed of one or more flexible fragments and sidechains, which are on the rigid part and near the mutated residue. We used dihedral angles as independent variables to describe the molecular conformation of the flexible part by fixing bond

lengths and bond angles to the standard values of ECEPP/2.[10,11] Each N-terminus of the flexible fragments has six degrees of freedom for translation and rotation relative to the rigid part.

ENERGY FUNCTION

The energy function of the system was represented as

$$E = \lambda_{nb} E_{nb} + \lambda_{tor} E_{tor} + \lambda_{vw} E_{vw} + \lambda_{lp} E_{lp} + \lambda_{\omega} E_{\omega} \quad (1)$$

where E_{nb} is the nonbonded interaction that is the summation of Lennard-Jones and electrostatic potentials over all atom pairs and E_{tor} is the intrinsic torsional potential around rotatable bonds. The parameters of ECEPP/2 were used for E_{nb} and E_{tor}. The term E_{vw} represents soft repulsion specifically introduced to allow atom overlapping, and it was used in early stages of simulated annealing instead of E_{nb} mentioned later. The function form of E_{vw} was

$$E_{vw} = \sum_{i<j} W_{ij}\left\{r_{ij}^2 - (s_i + s_j)^2\right\}^2$$
$$W_{ij} = \begin{cases} 0 & r_{ij} > s_i + s_j \\ 1 & r_{ij} \le s_i + s_j \end{cases} \quad (2)$$

where r_{ij} is the distance between an atom i and an atom j, and s_i is the van der Waals radius of the atom i. The values of the radii were the same as those used in the distance geometry calculation.[21] The energy E_{vw} is calculated faster than E_{nb} because the function form is simple and a cutoff algorithm for short-range force is effectively used. The term E_{lp} in eq. (1) is an artificial constraint to anchor the N- and C-termini of the flexible fragment to the rigid part (details provided later). Another artificial constraint, E_{ω}, restrains omega angles to the trans (or cis) position during the Monte Carlo simulation even at high temperatures. The function form of E_{ω} was

$$E_{\omega} = \sum_a \left\{\left(\frac{\theta_a}{\pi}\right)^2 - 1\right\}^2 \quad (-\pi < \theta_a \le \pi) \quad (3)$$

where θ_a is the ath omega angle. This form has been used for the angle constraint in the distance geometry calculation.[22] The parameters, $\{\lambda\}$, are weights used to modulate the contribution of each energy term to the total energy E.

We set the boundaries between the flexible fragment and the rigid part at the position of peptide bonds. To define the omega angles at these peptide bonds and to evaluate their torsional potentials, we attached pseudoatoms C_{α}^* and C* at the N-terminus of the flexible fragment and N* and C_{α}^* at the C-terminus (see Fig. 1). The pseudoatoms were excluded from the nonbonded energy calculation. The N- and C-termini of the flexible fragment should be anchored to the correct positions in the rigid part such that the main chain connects normally. Thus, we also put pseudoatom N* at the boundary of the N-terminal side of the rigid part and C* at that of the C-terminal side. The expression of E_{lp} was

$$E_{lp} = \sum_i |\mathbf{r}_i - \mathbf{r}_{i*}|^2 \quad (4)$$

where $\mathbf{r}_i$ is a vector representing the position of the anchored atom i (i.e., C_{α}, C, and N at the boundaries of both the N- and C-terminal sides) and $\mathbf{r}_{i*}$ is a vector representing the position of the

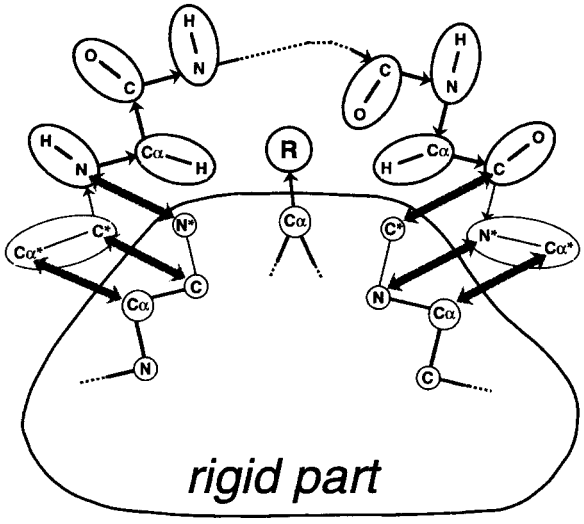

FIGURE 1. Diagram of the molecular system used in ESAP. This is composed of a rigid part, one flexible fragment, and one rotatable sidechain (R) on the rigid part. The flexible fragment is represented by a tree structure composed of units (ellipsoids) connected by rotatable bonds (arrows).[7] The rigid part is regarded as a large unit within which the energy calculation for atom pairs is omitted. Sidechains in the flexible fragment are omitted for simplicity. Thin lines with and without arrowheads represent hypothetical bonds connecting pseudoatoms (denoted by asterisks at the atom names). Pseudoatoms were introduced to anchor the flexible fragment to the rigid part. Thick hatched arrows show artificial attraction forces for the anchoring.

pseudoatom corresponding to the atom i. Generally, E_{lp} consists of loop constraints to form cyclic structures of chemical bonds, such as disulfide bridges between cysteine residues.

The computation time to calculate E is proportional to the number of atom pairs because the most time-consuming process is the calculation of atomic pairwise interactions (E_{nb} and E_{vw}). In the ESAP method, however, it is possible to omit the calculation for atom pairs within the rigid part. When the system consists of a rigid part of M atoms and a flexible part of N atoms (usually $M \gg N$), the order of the number of atom pairs taken in the energy calculation is reduced from $(M + N)^2$ to MN. Thus the introduction of the rigid part drastically reduces the computation time.

MONTE CARLO SIMULATION

Efficient conformation sampling was archived by using scaled collective variables[4] to modify the Metropolis Monte Carlo method.[23] The Hessian was calculated every 1000 steps of the simulation only if the acceptance ratio during the last 1000 steps was below 0.1. Empirical values of other parameters to prescribe this efficient way of sampling were the same as those in a previous report.[1] We used the FEDER/2 program, which is completely independent from the previous program.[1] In FEDER/2, the algorithm of the rapid Hessian calculation for multiple chains was used.[24,25] A detailed description of the FEDER/2 program will be published elsewhere.[13]

RESIDUE MUTATIONS IN T4 LYSOZYME

We focused on mutations at the 86th residue in T4 lysozyme. This residue is proline in the wild type. However, the X-ray structure of the Ala mutant was selected as the structure of the rigid part because alanine is a neutral amino acid without notable features. The performance of the improved ESAP should be tested for such a simple exercise to model structural changes from the Ala mutant to the other mutants, which include the wild type regarded as the Pro mutant.

The structure of the rigid part was fixed to the X-ray structure of the Ala mutant, which was regularized with the bond lengths and bond angles resulting from ECEPP/2. The dihedral angles of methyl and hydroxyl groups were determined with energy minimization, in which the other dihedral angles were fixed to the regularized positions. The first to the 70th and the 134th to the 164th residues in the rigid part were excluded from the energy calculation in the Monte Carlo simulation. Those residues are 20 Å or more away from Ala86 in the X-ray structure. No other cutoff operation was done.

The peptide fragment from Lys85 to Val87 and the sidechain of Asp89 was defined as the flexible part. The flexible fragment included the dihedral angles from the omega of Leu84 to the omega of Val87 along the main chain. In the X-ray structure of the Ala mutant, the geometrical centers of Leu84, Lys85, Val87, and Asp89, are within 6 Å away from that of Ala86. Leu84, which is buried in the hydrophobic core,[19] was assumed to have little conformational change as a result of the residue mutation. Thus Leu84 was not included in the flexible part. We did calculations for 10 kinds of flexible fragments, in which the 86th residues were Ala, Asp, Asp−, Gly, His, Leu, Arg, Ser, Pro, and Ile. Here Asp and Asp− denote protonated (neutral) and unprotonated (negatively charged) aspartic acid residues, respectively. Note that we used one unique rigid part taken from the Ala mutant to model the structures of these 10 kinds of flexible parts.

EXTENDED SIMULATED ANNEALING PROCESS

The ESAP consisted of seven stages of Monte Carlo simulation. The weights $\{\lambda\}$ of each energy term in the energy function E [eq. (1)] were varied, as shown in Table 1. The soft repulsion E_{vw} was used instead of E_{nb} in the first and second stages of the simulation. The value of λ_{lp} at the final stage was chosen so that the root mean square fluctuation of the atoms anchored to the rigid part became about 0.1 Å. The extended chain was taken as the initial conformation of the flexible fragment, and it was put at a position that was 12 Å apart from C_α of Ala86 in the X-ray structure. The orientation of the extended chain was set arbitrarily but without atom overlapping between the chain and the rigid part. All simulations started from this unique extended conformation with different series of random numbers. The first to sixth stages were used for annealing the system, whereby the lowest-energy conformation in each stage was used as the initial conformation of the next stage. The seventh stage was used for conformation sampling at equilibrium. The initial conformation of the seventh stage was the last accepted conformation of the sixth stage. The values of $\{\lambda\}$ at the sixth stage were completely the same as those at the seventh.

TABLE I.
Parameters of Extended Simulated Annealing Process (ESAP).

Stage	1	2	3	4	5	6	7
λ_{nb}	0	0	1	1	1	1	1
λ_{tor}	0	0	1	1	1	1	1
λ_{vw}	0.01	0.01	0	0	0	0	0
λ_{lp}	1	1	10	10	20	40	40
λ_{ω}	1000	1000	1000	200	40	0	0
Temperature	30,000	30,000 −4000[a]	30,000	6000	1200	300	300
No. of steps	20,000	20,000	50,000 (80,000)[b]	20,000	20,000	20,000	40,000

[a] Temperature was exponentially decreased from 30,000 to 4000 K by following the function form 30,000 * exp(−0.0001 * step).
[b] Calculation was continued to 80,000 steps when the energy value did not reach a plateau within 50,000 steps.

To improve the efficiency of the computation of ESAP, we selected the initial structure of the third stage by monitoring its energy value. When the initial energy was higher than 10^8 kcal/mol, the Monte Carlo simulation was quit and another simulation was restarted from the first stage. In other words, the first two stages were repeated until a favorable initial structure with an energy lower than 10^8 kcal/mol was obtained. We did eight independent simulations up to the last stage for each mutant.

ANALYSIS OF THE RESULTS

We defined the internal energy, U, of the system as the mean energy during the last stage of ESAP at 300 K, as described in the previous article. The model structure of the flexible part was also defined as the averaged conformation over the accepted conformations during the last stage.[1] In this article this model structure is called the ESAP structure.

We compared the ESAP structure with the X-ray structure for each mutant. First, the rigid part was superimposed on the corresponding part taken from the X-ray structure of the mutant concerned. Next, the root mean square deviation (rmsd) between the ESAP structure and the X-ray structure corresponding to the flexible part was calculated. We did not directly superimpose the ESAP structure on the X-ray structure, which was the correct answer of the predicted structure. We calculated two kinds of rmsd values: the rmsd value averaged over all heavy atoms and that averaged over main-chain atoms (N, C_α, and C). When the sidechain included topologically equivalent atoms, such as $O_{\delta 1}$ and $O_{\delta 2}$ of Asp89, the coordinates of the two atoms in the ESAP structure were exchanged and another rmsd was calculated. Then a smaller value between the two rmsd's was selected as the correct one.

The atomic coordinates of the T4 lysozyme mutants were taken from the Brookhaven Protein Data Bank[26]; the entry codes were 3LZM, which is the wild type (the 86th residue is Pro), 1L25 (Ala), 1L27 (Asp), 1L28 (Gly), 1L29 (His), 1L30 (Leu), 1L31 (Arg), and 1L32 (Ser). All computations were done on a MIPS RC6280 or RS3230 (Kubota Computer Inc.).

Results

EXTENDED SIMULATED ANNEALING PROCESS

For each mutant, we did about 10 simulations of the first two stages (Table II) to obtain eight conformations with energy values lower than 10^8 kcal/mol as the initial structure of the third stage. Eight energy profiles at the third stage for the Leu mutant are shown in Figure 2a. Six profiles quickly reached plateaus within 4000 steps. However, one profile (long dashed line) did not reach a plateau within 50,000 steps. The short dashed profile, which started at a higher energy value than 10^8 kcal/mol, is shown for comparison. Both the dashed profiles were continued up to 80,000 steps. The long dashed profile reached a plateau (data not shown) and the simulation of the next stage was continued, but the short dashed profile did not reach a plateau. Thus the threshold value of 10^8 kcal/mol was appropriate for these simulations. In the latter stages, there were no large energy differences observed among the eight profiles (data not shown).

TABLE II.
Statistics of ESAP Structures of T4 Lysozyme Mutants at the 86th Residue.

Residue Name	rmsd[a] (Å)	N_{vw}[c]	rmsd[d] (Å)	min rmsd[e] (Å)	ΔU[f] (kcal / mol)
Ala	0.177[b]	12	0.267	0.267	0.
Asp	0.193	10	0.449	0.438	1.6
Asp−		9	0.962	0.814	1.2
Asp− ($\varepsilon = 10$)[g]		↑	0.766	0.527	0.1
Asp− (all)[h]		↑	0.924	0.826	1.1
Gly	0.212	9	0.309	0.299	0.6
His	0.379	13	0.742	0.422	0.9
Leu	0.193	13	0.354	0.347	0.2
Arg	0.370	12	0.791 (0.389)[i]	0.791	0.
Ser	0.223	10	0.412	0.412	0.
Pro	0.273	16	0.656	0.644	2.3

[a] Root mean square deviation (rmsd) between the rigid part used for ESAP and the corresponding part taken from the X-ray structure for each mutant. All rmsd values in this table were calculated for all heavy atoms.
[b] The rmsd value represents the accuracy of regularization because the rigid part was taken from the X-ray structure of the Ala mutant (details described in text).
[c] Number of trials of the first two stages of ESAP required to obtain eight structures with energy values lower than 10^8 kcal / mol for the initial structure of the third stage.
[d] The rmsd value between the ESAP structure with the lowest internal energy U and the X-ray structure.
[e] Minimum rmsd value among eight ESAP structures.
[f] Difference between U of the structure with the minimum rmsd and the lowest U.
[g] ESAP structures of the Asp mutant with the charged Asp86 using a dielectric constant, ε, of 10. For all other mutants, $\varepsilon = 2$ was used. The symbol ↑ in N_{vw} means that the initial structures of the third stage were the same as those above.
[h] ESAP structures of the Asp mutant with charged Asp86 without the cutoff of electrostatic interaction.
[i] The rmsd value except for atoms with temperature factors larger than 50 in the X-ray structure of the Arg mutant.[19]

The energy profiles for each mutant reached plateaus in each stage of ESAP.

THE INTERNAL ENERGY AND RMSD

We compared each of eight ESAP structures with the corresponding X-ray structure for each mutant. Figure 3 shows the correlation between the rmsd and the internal energy U. We found that there was a tendency for the structures with a lower U to have a smaller rmsd. For Ala, Arg, and Ser mutants, the ESAP structure with the lowest U gave the smallest rmsd among the eight structures (Table II). For the other mutants, the rmsd with the lowest U was almost as small as the smallest rmsd. For any mutant, the root mean square fluctuation of the total energy E in the last stage was 2.0 to 2.5 kcal/mol, which was smaller than the differences of U among the ESAP structures. There was no clear correlation between the energy fluctuation and the rmsd (data not shown). These results show that U is a good criterion by which to select a correct model among the ESAP structures.

For five mutants (Ala, Asp, Gly, Leu, and Ser), the structure with the lowest U agreed with the X-ray structure within 0.5 Å rmsd for all heavy atoms (Table II). For the His and Arg mutants, the ESAP structures gave larger rmsd values than those for the other mutants (Figs. 3d and 3f). The ESAP structure with the lowest U had about 0.7 Å rmsd. The X-ray structure of the Arg mutant, however, included atoms with large temperature factors, which indicates that these atoms have large fluctuations. The rmsd values, except for atoms with temperature factors larger than 50, were as small as that of the other mutants (Fig. 3f). The excluded atoms from the rmsd calculation were those from C_δ to the end of the sidechain of Arg86 and N_ζ of Lys85. For the His mutant, the temperature factors of the atoms in the imidazole ring were also large (46 to 52). Thus the rmsd value with the lowest U was large. The result for the Pro mutant shows an example in which the size of the flexible part was too small (discussed later).

Except for the Pro mutant, the main-chain ESAP structure with the lowest U agreed with the main-chain X-ray structure within 0.2 Å rmsd (Fig. 3). This rmsd value was thought to be the resolution limit because the structure of the rigid part taken from the Ala mutant differed from the X-ray struc-

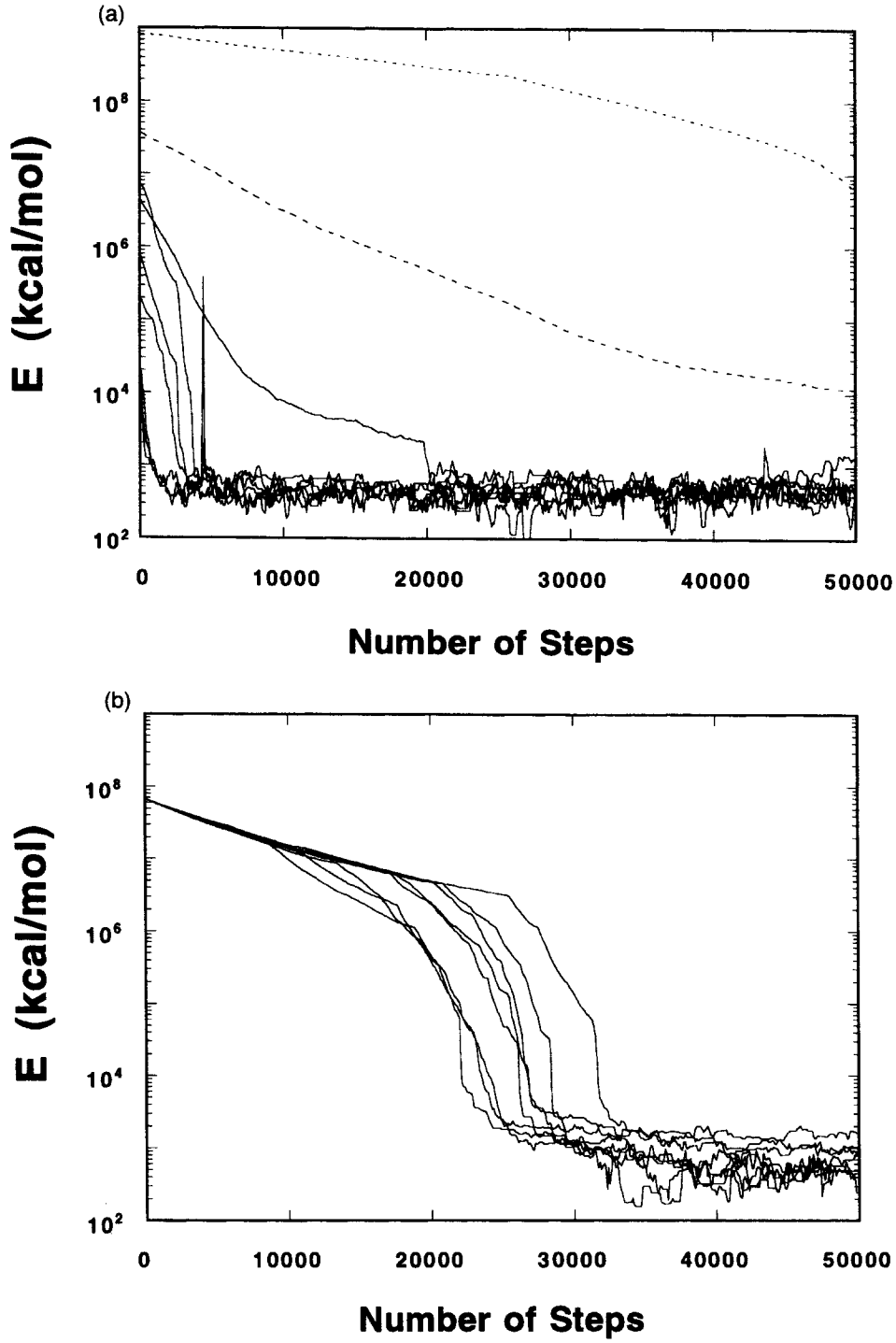

FIGURE 2. Energy profiles for the Leu mutant at the third stage of ESAP. (a) The simulation started from the structures obtained after the first and second stages of simulated annealing using only soft repulsion and (b) from the extended structure. The temperature used in the simulation was 30,000 K. Energy values averaged over every 100 steps are plotted. The dashed lines in Fig. 2a were continued up to 80,000 steps (details described in text).

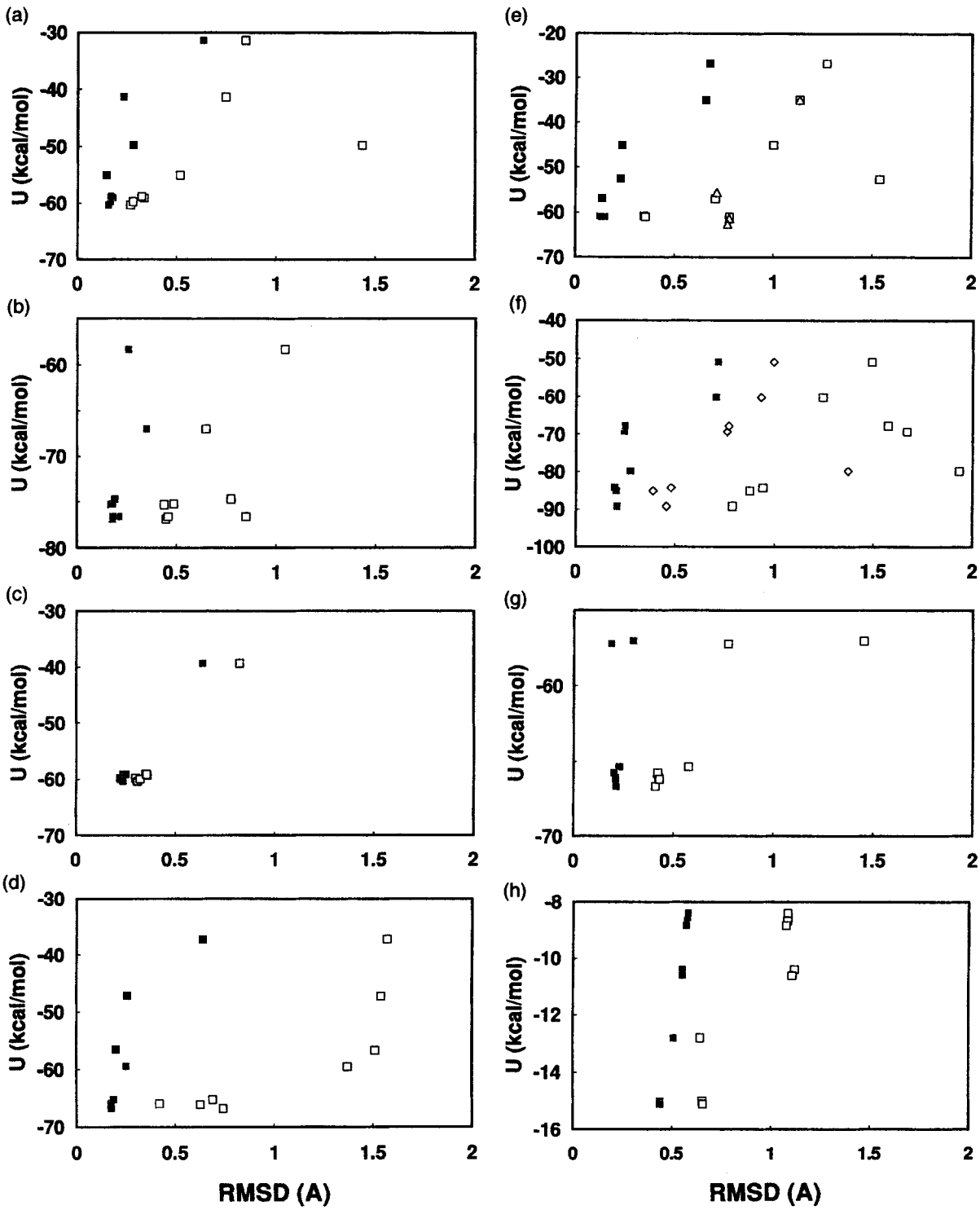

FIGURE 3. The correlation between rmsd and the internal energy U of ESAP structures for T4 lysozyme mutants. The 86th residue was mutated to Ala (a), Asp (b), Gly (c), His (d), Leu (e), Arg (f), Ser (g), and Pro (h). For each mutant, two kinds of rmsd between the ESAP structures and the X-ray structure were plotted. One was the rmsd for all heavy atoms (□) and the other was that for the main-chain atoms (■). In Fig. 3e the symbol Δ represents the ESAP structure obtained without the first and second stages of the simulation corresponding to Fig. 2b. The rmsd was calculated for all heavy atoms. In Fig. 3f the symbol ◇ denotes the rmsd calculated from only atoms with temperature factors less than 50 in the X-ray structure.[17]

ture of each mutant by the same degree of rmsd (Table II). In other words, the main-chain structure was completely predictable based on the framework of the flexible part used.

IMPROVEMENT WITH THE SOFT REPULSION POTENTIAL

As the atomic pairwise interaction, we used only the soft repulsion potential E_{vw} in the first and second stages and only E_{nb} from the third stage (Table I). To clarify the significance of the first and second stages, other annealing processes started from the third stage as a control. In this article ESAP simulations from the first stage are called ESAP with E_{vw} and those from the third stage are called ESAP without E_{vw}. Figure 2b shows eight energy profiles of ESAP without E_{vw} for the Leu mutant at 30,000 K. These profiles represent the starting stage of ESAP without E_{vw} but correspond to the third stage of ESAP with E_{vw}. Since these eight simulations started from the same extended conformation as that used in ESAP with E_{vw}, the profiles agreed with each other at the beginning of the simulation. All profiles reached a plateau within 50,000 steps. The energy level of the plateau is the same as that shown in Figure 2a. However, the number of steps to reach the plateau is the same as that shown in Figure 2a. Then we obtained eight ESAP structures without E_{vw} for the Leu mutant.

Figure 3e shows the correlation between the rmsd and U for four ESAP structures without E_{vw} together with the eight ESAP structures with E_{vw}. The other four ESAP structures without E_{vw}, which were not in the range of the figure, were trapped at local minima with high U. The ESAP structures with E_{vw} had smaller rmsd values than those without E_{vw}, although the level of the lowest U was nearly the same. The required central processing unit (CPU) time for one step of the Monte Carlo simulation using E_{vw} was one third of that using E_{nb}. Thus the soft repulsion used in the first two stages was effective for obtaining ESAP structures that were close to the X-ray structure.

EFFECT OF A CHARGE STATE

For the Asp mutant, we did ESAP simulations for two kinds of charge states in which the sidechain carboxyl group was either neutral (Asp) or negatively charged (Asp$-$). The ESAP structures of the neutral form were better than those of the charged form for both the degree of the correlation and the small value of the rmsd (Fig. 4a). A major structural difference between the two groups of ESAP structures was found at the sidechain of Lys85 (Fig. 4b). In the ESAP structures of the neutral form and the X-ray structure, Lys85 formed an ion pair with Asp89. In the ESAP structures of the charged form, however, the amino group of Lys85, which was located in the middle position between the carboxyl groups of Asp86 and Asp89, formed ion pairs with both of them. Thus the charged form had a lower value of U than the neutral form. Note that the difference of the absolute value of U did not represent the reality of the charged state but reflected the difference of the electrostatic environment.

To examine the effect of the difference of the electrostatic environment on the ESAP structure, other ESAP simulations differing in a dielectric constant, ε, were done from the third stage for the charged form (Table II). Besides ε of 2.0, which is the normal value in ECEPP/2, ε of 10.0 was used because the flexible part treated in this study was located on the protein surface. The ESAP structures of the charged form at $\varepsilon = 10$ were as good as those of the neutral form (Fig. 4).

Furthermore, to validate the cutoff of electrostatic interaction, the ESAP simulation without the cutoff was done for the charged form at $\varepsilon = 2$. The ESAP structure obtained was almost the same as those with the cutoff (Fig. 4a). Thus the cumulative effect of long-range electrostatic contributions can be ignored in the current simulation.

MODELING OF THE ISOLEUCINE MUTANT

We also obtained eight ESAP structures for the Ile mutant, for which the X-ray structure has not yet been reported. The internal energies of the ESAP structures were -57.4, -57.1, -55.3, -54.5, -54.2, -53.9, -24.6, and -14.1 kcal/mol in ascending order. To examine the structural differences among the eight structures, we calculated pairwise rmsd for all 28 pairs of the structures. Next, we did principal component analysis of an 8×8 matrix whose off-diagonal elements were the rmsd values (the diagonal ones, which represent the rmsd of a structure with itself, were zero). Figure 5 is a three-dimensional plot of the largest principal value of the rmsd, the second largest value, and U of the eight structures. The two ESAP structures with the lowest and the second lowest values of U were close to each other. The rmsd value between these two was 0.005 Å. A tendency for ESAP structures with a lower U to

have a smaller rmsd was found. Thus we propose that these two ESAP structures are plausible models for the Ile mutant. Based on the rmsd values for other mutants (Fig. 3), the model structure with the lowest U should be accurate to an rmsd of 0.5 Å.

Table III shows the dihedral angle values calculated from the atom positions of the ESAP struc-

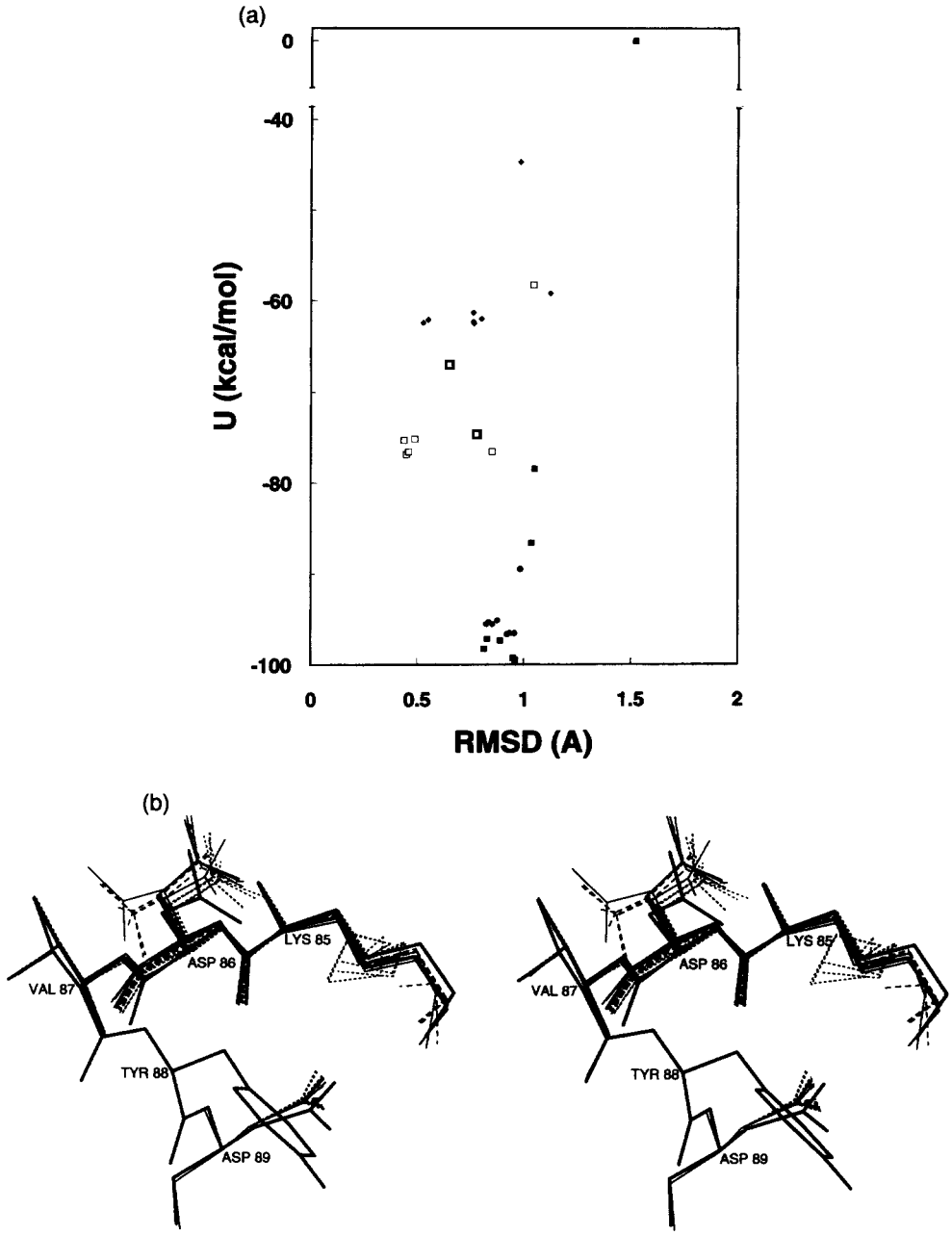

FIGURE 4. (a) The correlation between the rmsd and the internal energy U for the Asp mutant. (b) Stereo drawings of the ESAP structures. In Figs. 5a and b, the ESAP structures with neutral Asp86 are denoted by the symbol □ and thin solid lines, those with charged Asp86 at $\varepsilon = 2$ by ■ and short dashed lines, and those with charged Asp86 at $\varepsilon = 10$ by ♦ and long dashed lines. In Fig. 5a the symbol ● denotes the ESAP structures with charged Asp86 without the cutoff of electrostatic interaction at $\varepsilon = 2$. In Fig. 5f the five ESAP structures with the lowest to the fifth lowest U are displayed and the thick solid line is the X-ray structure.[17] The drawing was made with the program MolSkop (JEOL Ltd.).

ture with the lowest U. The distribution of the dihedral angle values in the last stage was a single peak for all angles except the methyl groups. The root mean square fluctuations of these values were 3 to 8 degrees, except the χ_2's of Lys85 and Asp89 (13 and 15 degrees, respectively). In other words, the sidechain rotamer was not found.

Discussion

In this article, we have used *ab initio* methods to predict the structure of single amino acid mutations at the 86th residue of T4 lysozyme. We did not refer to the structure data of the flexible part before the mutation, but we used the extended conformation as the initial structure. In multidimensional space, however, structures are often trapped at local minima near the initial structure during simulated annealing as well as simple minimization. To attain the correct structure, the anchored atoms at N- and C-termini of the flexible fragment need to overlap atoms in the rigid part and sometimes have to pass through them. Thus we used E_{vw} for the atomic pairwise interaction in the first and second stages to obtain a favorable initial structure for the third stage. Two atoms that interact only by E_{vw} could overlap during Monte Carlo simulation at high temperatures because the function value of E_{vw} is finite even at $r_{ij} = 0$ [eq. (2)]. The soft repulsion was effective for obtaining correct ESAP structures. A truncated Lennard-Jones potential, which also has a finite value at $r_{ij} = 0$, has been used in packing optimization of sidechain conformations to allow for atom overlapping.[16,17]

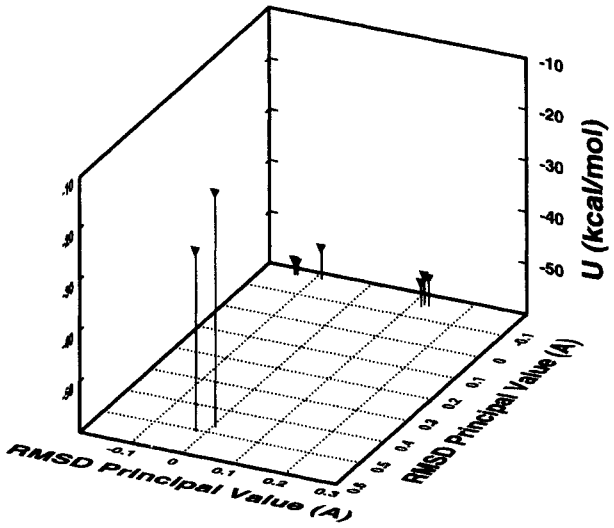

FIGURE 5. Three-dimensional representation of the internal energy U and the relative structural differences among ESAP structures for the Ile mutant. The structural differences are represented by the first and second largest principal values of the rmsd among the structures.

The initial structure of the third stage may have an extremely high energy value because E_{nb} was evaluated for the structure obtained in the second stage, in which E_{vw} was used instead of E_{nb}. Annealing processes from structures with such high energies are time consuming. However, the structures selected by the energy threshold of 10^8 kcal/mol were favorable because the energy profile of the third stage quickly reached a plateau. In the subsequent long simulation at equilibrium, a low-energy structure was explored and used as the initial structure for the next stage. Figure 2a shows that the six profiles quickly reached the plateau within 4000 steps. In such profiles the simulation of 50,000 steps may be excessive. In these six simulations, the initial structures for the next stage were obtained at 10,198, 11,773, 21,446, 21,910, 36,572, and 36,577 steps. To save computation time, the simulation may be quit when it proceeds for a given number of steps at the plateau of the energy level. For a structure with a higher energy than the threshold (the short dashed line in Fig. 2a), another

TABLE III.
Dihedral Angle Values of the Ile Mutant Model.

Residue	ϕ	ψ	ω	χ_1	χ_2	χ_3	χ_4
Leu84			−176.7				
Lys85	−56.3	−44.2	−174.3	−175.6	−166.9	−171.8	−61.5
Ile86	−63.4	−29.8	177.3	−69.2	160.8	nd[a]	nd
Val87	−76.8	−53.6	−177.6	171.3	nd	nd	
Asp89				−70.9	170.2		

[a] nd = not determined. Dihedral angles of methyl groups in the ESAP structure were not determined because they were rotating at 300 K.

annealing process from a higher temperature may be effective.

In the ESAP structures obtained, the main-chain structures were determined with higher accuracy than those of the sidechains. The ESAP structure with the lowest internal energy agreed with the X-ray structure within 0.5 Å rmsd for all heavy atoms and within 0.2 Å for the main-chain atoms. Even ESAP structures with a high U had a small rmsd for the main-chain atoms. (Figs. 3a, 3b, 3d, and 3g). This result shows that the differences observed for U were mainly derived from the differences in the sidechain conformations. The main-chain conformation of a short flexible fragment is controlled by geometrical constraints arising from the surface shape of the rigid part and from the anchoring of both termini of the fragment to the rigid part. In the ESAP method, moreover, the dihedral angles are variables to describe the conformation of the fragment (i.e., bond lengths and bond angles are fixed). Thus the constraint arising from the anchoring exerts a straightforward effect on the main-chain conformation.

For the Pro mutant, even the main-chain ESAP structure had a larger deviation (about 0.5 Å) from the X-ray structure than those of the other mutants (Fig. 3h). In the X-ray structure, Ala86 is located at an α-helix, and a hydrogen bond between the N of Ala86 and the O of Ala82 exists. It is easy to suppose that, for the Pro mutant, the conformational changes may occur in the N-terminal side of Pro86 because the O of Ala82 cannot bind to the N of Pro86. It has been reported that a conformational change from the Pro mutant to the Ala mutant occurred locally in residues 81 to 83,[19] although the overall rmsd between the rigid part taken from the Ala mutant and the Pro mutant (0.273 Å) is not significant in Table II. Thus, for the modeling of the Pro mutant from the Ala mutant, the reasonable size of the flexible part was larger than that used, which included Lys85 to Val87 and the sidechain of Asp89. In the preliminary calculation using the large flexible part, which included Asn81 to Val87, no ESAP structure with a smaller rmsd was obtained at the same simulation step. It was difficult to fold the flexible fragment of seven residues from the extended conformation to an α-helix. For such a large flexible part, it may be necessary to use the structure before the mutation as the initial conformation.

For a large flexible part, more refined definition of the ESAP structure may be required. In the ESAP structure of the Ile mutant shown in Table III, the structure fluctuation was analyzed. However, the root mean square fluctuation of the dihedral angle value is believed to be smaller than real one because the motion of all atoms in the rigid part was freezed. Therefore, the sidechain rotamer was seldom found. For a larger flexible part, the distribution may be double peaks. Then cluster analysis of modeled structures and a probability map of atom positions will be needed.[27]

Finally, we discuss the effect of the charge state and the dielectric constant on the ESAP structure of the Asp mutant. The simulation result that both ESAP structures of the neutral form and the charged form at $\varepsilon = 10$ were close to the X-ray structure suggests two possibilities. One is that the 86th aspartic acid residue is actually protonated, and the other is that its negative charge is shielded by water molecules or counterions because T4 lysozyme crystals were equilibrated with a solution including at least 1 M K^+ and 1.5 M Na^+.[28] A dielectric constant was an adjustable parameter to the surrounding electrostatic environment.[29] It does not matter that a substance with ε of 10 really exists, although the value itself was an apparent ε of hydrated water on the surface of nonionized silica gels.[30] Electrostatic interaction at $\varepsilon = 2$ may be too strong on the protein surface. The energy parameters of ECEPP/2 were based on nonbonded interactions in a vacuum. In this article neither water molecules nor hydration were explicitly considered. The solvent effect, however, is implicitly included in the structure of the rigid part that is determined by X-ray crystallography. Thus the ESAP structures were close to the X-ray structure for each mutant. As shown in the example of the Asp mutant, a better ESAP structure of charged sidechains on the protein surface may be obtained in a high-dielectric environment.

Conclusion

We have implemented ESAP so that it can be used for general macromolecules using FEDER/2, which provides fast analytical calculation of the Hessian. Using the improved ESAP, we modeled the structural changes resulting from eight kinds of mutation at the 86th residue in T4 lysozyme. The soft repulsion between the van der Waals spheres was effectively used to improve the accuracy of the modeling. The internal energy was a good criterion for selecting a correct model among ESAP structures. For each mutant, we obtained model structures that were close to the X-ray

structure. The rmsd's from the X-ray structure were 0.3 to 0.8 Å for all heavy atoms and about 0.2 Å for the main-chain atoms. We also modeled the structure of the Ile mutant for which the X-ray structure has not yet been reported. ESAP can be used to model structural changes caused by single-residue mutation in proteins.

Acknowledgment

The authors thank Dr. Jean-Yves Trosset of Baker Laboratory, Cornell University, for helpful suggestions regarding Monte Carlo simulation.

References

1. J. Higo, V. Collura, and J. Garnier, *Biopolymers*, **32**, 33 (1992).
2. J. F. Gibrat, J. Higo, V. Collura, and J. Garnier, *Immunomethods*, **1**, 107 (1993).
3. V. Collura, J. Higo, and J. Garnier, *Protein Sci.*, **2**, 1502 (1993).
4. T. Noguti and N. Gō, *Biopolymers*, **24**, 527 (1985).
5. G. Vasmatzis, R. Brower, and C. DeLisi, *Biopolymers*, **34**, 1669 (1994).
6. T. Noguti and N. Gō, *J. Phys. Soc. Japan*, **52**, 3685 (1983).
7. H. Abe, W. Braun, T. Noguti, and N. Gō, *Comp. Chem.*, **8**, 239 (1984).
8. H. Wako and N. Gō, *J. Comp. Chem.*, **8**, 625 (1987).
9. T. Schaumann, W. Braun, and K. Wüthrich, *Biopolymers*, **29**, 679 (1990).
10. F. A. Momany, R. F. McGuire, A. W. Burgess, and H. A. Scheraga, *J. Phys. Chem.*, **79**, 2361 (1975).
11. G. Némethy, M. S. Pottle, and H. A. Scheraga, *J. Phys. Chem.*, **87**, 1883 (1983).
12. S. Endo, H. Wako, K. Nagayama, and N. Gō, *NATO ASI Series*, **A225**, 233 (1991).
13. H. Wako, S. Endo, K. Nagayama, and N. Gō, *Comput. Phys. Commun.*, in press (1995).
14. B. Robson and E. Platt, *J. Mol. Biol.*, **188**, 259 (1986).
15. R. M. Kini and H. J. Evans, *J. Biomolec. Str. Dyn.*, **9**, 475 (1991).
16. C. Lee and S. Subbiah, *J. Mol. Biol.*, **217**, 373 (1991).
17. L. Holm and C. Sanders, *Proteins*, **14**, 213 (1992).
18. A. P. Heiner, J. C. Berendsen, and W. F. van Gunsrteren, *Protein Eng.*, **6**, 397 (1993).
19. T. Alber, J. A. Bell, S. Dao-Pin, H. Nicholson, J. A. Wozniak, S. Cook, and B. W. Matthews, *Science*, **239**, 631 (1988).
20. L. H. Weaver and B. W. Matthews, *J. Mol. Biol.*, **193**, 189 (1987).
21. W. Braun and N. Gō, *J. Mol. Biol.*, **186**, 611 (1985).
22. W. Braun, *Q. Rev. Biophys.*, **19**, 115 (1987).
23. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys.*, **21**, 1087 (1953).
24. W. Braun, S. Yoshioki, and N. Gō, *J. Phys. Soc. Japan*, **53**, 3269 (1984).
25. J. Higo, Y. Seno, and N. Gō, *J. Phys. Soc. Japan*, **54**, 4053 (1985).
26. F. C. Bernstein, T. F. Koetzle, G. J. B. Williams, E. F. Meyer, Jr., M. D. Brice, J. R. Rodgers. O. Kennard, T. Shimanouchi, and M. Tasumi, *J. Mol. Biol.*, **112**, 535 (1977).
27. W. L. DeLano and A. Brünger, *Proteins*, **20**, 105 (1994).
28. S. J. Remington, W. F. Andeson, J. Owen, L. F. Ten Eyke, C. T. Grainger, and B. W. Matthews, *J. Mol. Biol.*, **118**, 81 (1978).
29. T. Montcalm, W. Cui, H. Zhao, F. Guarnieri and S. R. Wilson, *J. Mol. Struc.*, **308**, 37 (1994).
30. T. Sakamoto, H. Nakamura, H. Uedaira, and A. Wada, *J. Phys. Chem.*, **93**, 357 (1989).